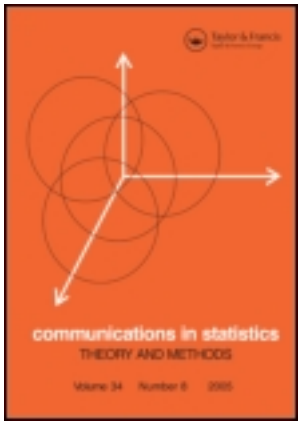

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

Handling Uncertainty in the Cost Effectiveness Healthcare Evaluations. A Review of Statistical Approaches

Vilelmine Carayanni [a]
[a] Department of Public Hygiene, Highest Technological Institute of Athens, Egaleo, Greece
Published online: 01 Apr 2009.

To cite this article: Vilelmine Carayanni (2009): Handling Uncertainty in the Cost Effectiveness Healthcare Evaluations. A Review of Statistical Approaches, Communications in Statistics - Theory and Methods, 38:8, 1224-1240

To link to this article: http://dx.doi.org/10.1080/03610920802393079

Communications in Statistics—Theory and Methods, 38: 1224–1240, 2009

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920802393079

Handling Uncertainty in the Cost Effectiveness Healthcare Evaluations. A Review of Statistical Approaches

VILELMINE CARAYANNI

Department of Public Hygiene, Highest Technological Institute of Athens, Egaleo, Greece

Healthcare economic evaluation is an analytical tool used with increasing frequency to assist decision making in the choice and financing of interventions and technologies in the healthcare system. The objective of this article is to analyze the different methods of handling with sampling uncertainty in healthcare cost effectiveness evaluations when patient level data are available. The aim of this article is to focus on the strengths and the weakness of each method in order to facilitate the tasks of those who must base their choices on studies of this kind.

Keywords Healthcare evaluation; Sampling uncertainty; Stochastic cost effectiveness analyses.

Mathematics Subject Classification Primary 92B15.

1. Introduction

Healthcare economic evaluations have been used with increasing frequency in recent years for the following reasons: (1) the population is aging; (2) the number and the type of the professionals in the health sector increase; (3) the medical techniques in every field develop; (4) the financial limitations impose the control of health expenses.

In particular, there is a considerable interest from health providers worldwide in assessing the cost effectiveness (CE) of new treatments. Pharmacoeconomic analyses are being used increasingly as the basis for reimbursement of the costs of new drugs. Reports of these analyses are often published in peer-reviewed journals. However, the analyses are complex and difficult to evaluate and very little guidance is given to researchers on exactly how the assessment of the implications of uncertainty should be done and how the results of the analysis should be presented. The problem is more serious for the stochastic evaluations (evaluations using exclusively patient level data). Although there is an increasing trend towards

Received March 16, 2007; Accepted August 6, 2008
Address correspondence to Vilelmine Carayanni, Department of Public Hygiene, Highest Technological Institute of Athens, 28 Saint Spyridonos Street, Egaleo 12210, Greece; E-mail: bilma@otenet.gr

conducting evaluations alongside clinical trials, only a small proportion of all economic evaluations (almost 6%) in the last decade have used exclusively patient–level data (Briggs, 2001). Nevertheless, the quality criteria laid down by clinical epidemiologists for clinicians consider randomized trials as the best way to offer estimates of treatment effects (efficacy).

On the other hand, pragmatic clinical trials are strongly recommended by health economists to increase external validity and generalisability of the results. So, it is considered very interesting to focus our study on handling the sampling uncertainty in the case of cost effectiveness (CE) stochastic evaluations, this type of evaluations being more frequently used by the researchers than other types of evaluation, such as cost utility and cost benefit analysis (Pritchard, 2004). In Sec. 2 the general model of evaluation and the different types of uncertainty that can arise in cost effectiveness analysis are briefly presented. In Sec. 3, the methods of handling the sampling uncertainty in stochastic cost effectiveness analysis are discussed when uncertainty is limited to the III quadrant of the CE plane (estimation of confidence intervals around Incremental Cost Effectiveness Ratios–ICER's) and when the uncertainty covers the more than one quadrant (alternative decision methods based on CE ratios). A supplementary distinction has been used for the methods according to: (1) the methods based on the assumption of normality; and (2) the nonparametric and the Bayesian methods. We will focus our analysis on the most widely used methods.[1] Section 4 offers a discussion of the issues raised in the article.

2. The General Model of Evaluation, the CE Plane and the Types of Uncertainty

2.1. *The General Model of Evaluation in Healthcare*

We define as general model of evaluation, concerning a subject of evaluation the quintuplet $[S, D,(\xi_i, E_i, >_i)_{i=1,\ldots,p}, X(\cdot), >_D]$ (Auray et al., 1991) as:

S is the total of situations to evaluate,

D is the observer,

p is the number of ξ_i criteria selected to describe the situations,

E_i is the total of the modalities of the criteria,

$\xi_i, >_i$ is a relation of preference complete, reflexive and transitive defined on E_i,

$X(\cdot)$ the descriptor of the S, that is an application of S to

$X(\cdot) = [X_1(\cdot), \ldots, X_p(\cdot)]$,

$E_1 x \cdots x E_p, >_D$ is the relation of preference on S constructed by D.

The classical methods (Table 1) for building the preference $> D$ are cost effectiveness analysis (monetary units/natural units), cost utility analysis (monetary units/healthy years), and cost benefit analysis (monetary units/monetary units). Table 2 presents a brief summary of real examples using cost effectiveness analysis in healthcare evaluations.

[1]Other methods that are rarely used, such as Bonferonni, Angular Transformations, and Bayesian (nonparametric) bootstrap, are not discussed in this article.

Table 1
The classical methods for building the preference $>D$

Nomenclatures	Techniques
Monetary criteria-ONE non monetary criterion	Cost effectiveness
Monetary criteria-and non monetary criteria	Cost utility (construction of a utility function)
ALL criteria are monetary	Cost benefit

To construct an evaluation model, the four following concepts will be used.

1. **Utility deviation between the situations s and s' in reference to a criterion.** Given two situations, s and s' of S, we name utility deviation in reference to ξ_i the quantity (s, s', i) defined by:

$$\Delta(s, s', i) = |u_i(s) - u_i(s')|, \tag{1}$$

where u_i is a cardinal utility function.

2. **Cost of s in reference to s'.** Given two situations, s and s' of S,

$$C(s, s') = \sum_{i \in X^-(s,s')} \Delta(s, s', i), \tag{2}$$

where $X^-(s, s')$ is the total of criteria for that s' is preferred than s.

3. **Benefit of s in reference to s'.** Given two situations, s and s' of S, is named utility deviation in reference to a criterion ξ_i the quantity (s, s', i) defined by:

$$B(s, s') = \sum_{i \in X^+(s,s')} \Delta(s, s', i), \tag{3}$$

where $X^+(s, s')$ is the total of criteria for that s is preferred than s'.

4. **Gains of s in reference to s'.** If we have two situations, s and s', is named gain of s in reference to s' the quantity $G(s, s')$ defined by:

$$G(s, s') = B(s, s') - C(s, s'). \tag{4}$$

2.2. *The Cost Effectiveness Aggregation and the Cost Effectiveness Plane*

There are three types of aggregation for the cost effectiveness analyses.

1. *Paretian aggregation.* A situation s is preferred globally than a situation s' if the monetary gain procured by the passage of a situation of reference $\hat{s}$ to s is better than the gain for s' and s is more effective than s':

$$s >_D s' \Leftrightarrow \widehat{G}(s) \geq \widehat{G}(s') \quad \text{and} \quad s \succ_{iE} s'. \tag{5}$$

In general, the $\widehat{G}(s)$ and $\widehat{G}(s')$ are negative gains. In this case, it is evident that the $>_D$ isn't generally a complete relation.

Table 2
Examples of using the cost effectiveness analysis in healthcare evaluations

Reference	Clinical field	Type of intervention	Effectiveness measure
Wykes et al. (2007)	Cognitive remediation therapy for schizophrenia	Rehabilitation	The proportion of patients improving their WAIS-III raw score
Sophonsritsuk et al. (2005)	In vitro fertilization	Treatment	The pregnancy rate
Mandelblatt et al. (2004)	Therapy of breast cancer	Screening and treatment	Life years saved
Pinkerton et al. (2000)	Prevention of perinatal HIV transmission	Primary prevention	The number of infections averted
Armstrong et al. (2000)	Treatment of influenza	Treatment	The number of successfully treated patients (with no adverse reaction)

2. *Cost minimization aggregation.* We suppose that all situations of S are equivalents about their effectiveness and we compare monetary gains, that is,

$$\forall s, s' \in S \quad s \succ_{iE} s' \quad \text{and} \quad s' \succ_{iE} s \longrightarrow s \approx_{IE} s'. \tag{6}$$

In this case, the $>_D$ is generally a complete relation.

3. *Incremental cost effectiveness aggregation and the cost effectiveness plane.* Suppose that we are comparing a new pharmaceutical therapy with an existing therapy (or control therapy) which represents the better ratio cost effectiveness. Additionally, suppose that we know the true mean cost μ_{CN} and the true health outcome μ_{EN} of the new therapy versus the existing therapy (μ_{CS} and μ_{ES}, respectively). O'Brien et al. (1994) identified four situations that can arise in relation to the incremental cost and health outcome of therapies:

1. $\mu_{CN} - \mu_{CS} < 0; \mu_{EN} - \mu_{ES} > 0;$
   situation of dominance for the new therapy. (7)
2. $\mu_{CN} - \mu_{CS} > 0; \mu_{EN} - \mu_{ES} < 0;$
   situation of dominance for the existing therapy. (8)
3. $\mu_{CN} - \mu_{CS} > 0; \mu_{EN} - \mu_{ES} > 0;$
   Trade off, according to the paretian aggregation. (9)
4. $\mu_{CN} - \mu_{CS} < 0; \mu_{EN} - \mu_{ES} < 0;$
   Trade off according to the paretian aggregation. (10)

The proposed model for Cases 3 and 4 is the aggregation incremental cost effectiveness which, by definition, is:

$$\text{ICER} = \frac{\mu_{CN} - \mu_{CS}}{\mu_{EN} - \mu_{ES}} = \frac{\mu_{\Delta C}}{\mu_{\Delta E}}. \tag{11}$$

These situations are equivalent to the four quadrants of the cost effectiveness plane, presented in Fig. 1.

If the new therapy is situated in the quadrant III, the choice of the decision maker depends on the maximum cost effectiveness ratio that he is inclined to pay in order to pass from the standard therapy to the new therapy, that is to say from the maximum sum that he is willing to pay to obtain a supplementary unit of effectiveness by using the new therapy. Let it be noted R_{III} is the value.

If the ICER is of lesser value than R_{III}, then the new therapy is considered as better than the standard therapy.

If the new therapy is situated in the quadrant IV, the choice of the decision maker depends on the minimum cost effectiveness ratio which he is inclined to gain to pass from the standard therapy to the new therapy, that is of the minimum sum he agrees to pay at the expense of losing a unit of effectiveness by using the new therapy. Let it be noted R_{IV} is the value.

If the ICER is superior to the ratio R_{IV}, then the new therapy is considered as if it were better than the old therapy.

It is evident that this rule supposes that we know the real values of the mean costs and effectiveness used in formula (11). In practice, these values are not always

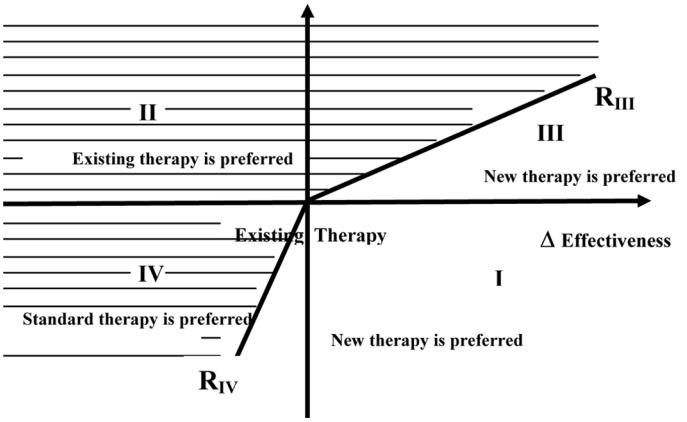

Figure 1. The cost effectiveness plane.

known but estimations of the real values can be obtained through the representative samples.

In fact, we work with the ratio

$$\overline{\overline{\text{ICER}}} = \frac{\bar{\bar{\mu}}_{CN} - \bar{\bar{\mu}}_{CS}}{\bar{\bar{\mu}}_{EN} - \bar{\bar{\mu}}_{ES}} = \frac{\bar{\bar{\mu}}_{\Delta C}}{\bar{\bar{\mu}}_{\Delta E}}, \tag{12}$$

where $\bar{\bar{x}}$ expresses the point estimation of parameter x.

In what follows, through abuse of the notion but to facilitate the flow of the writing, we note with the same fashion the parameter in search of estimation and its estimator.

The point estimation of the standard error in costs and the difference in effectiveness is given by:

$$\sqrt{\frac{s^2_{CN}}{n_N} + \frac{s^2_{CS}}{n_S}} \quad \text{and} \quad \sqrt{\frac{s^2_{EN}}{n_N} + \frac{s^2_{ES}}{n_S}} \tag{13}$$

where s^2_{CN}, s^2_{EN} are the estimated variances of the costs and the effects of the new therapy s^2_{CS}, s^2_{ES} are the estimated variances of the costs and the effects of the standard therapy for the samples sizes, n_N and n_S, respectively.

2.3. *Sources of Uncertainty*

The next crucial stage for a strength economic evaluation is to handle with the uncertainty.

We can relate (Table 3) the different approaches concerning the different sources of sampling uncertainty: similar taxonomies have been described by Briggs (2001), the US Panel on Cost Effectiveness (Gold et al., 1996; Spiegelhalter et al., 2000).

The first source of uncertainty is the methodology used. This uncertainty concerns the analytical methods used and the methods selected to value the resource and health consequences. The use of a reference case of methods in combination with one-way sensitivity analysis has great appeal in cost-effectiveness analysis, where results of a study only have meaning in comparison to the results of other studies.

Table 3
Methods for handling uncertainty in stochastic analyses

Sources of uncertainty	Handling uncertainty
Methodology	Reference case methods/ deterministic sensitivity analysis
Setting of the study	Probabilistic sensitivity analysis
Sampling variations (second order uncertainty)	Statistical analysis
Surrogate outcomes/ short term observation	Modelling

The second source of uncertainty is the setting of the study because that can be very specific (i.e., a study that was undertaken in a university hospital can include both higher costs and outcomes than a typical hospital). Consequently, this event will affect the transferability/generalizability of the results. The use of sensitivity analysis can be useful for handling this uncertainty. Probabilistic sensitivity analysis is the more appropriate method if our sample is representative, because we have parameters that could in principle be estimated from sample data.

The third source of uncertainty concerns the cases where clinical trials collect information only on surrogate points (i.e., mm Hg blood pressure reduction) and are observed in the short term rather than on the ultimate health outcomes of interest (i.e., death, illness). Different types of modelling (i.e., duration models, longitudinal models, etc.) can be used for these purposes.

Finally, it is evident that, for stochastic CE analyses, the sampling variation constitutes a source of uncertainty. For handling this uncertainty different statistical methods have to be used that we will see in Sec. 3.

3. The Methods for Handling the Sampling Uncertainty

3.1. *The Methods of Construction of Confidence Intervals Around ICER's*

Here, the methods to handle the sampling uncertainty are discussed when uncertainty is limited to the III quadrant of the CE plane A supplementary distinction has been used for the methods according to: A. The methods based on the assumption of normality. B. The nonparametric and the Bayesian methods.

3.1.1. *The Method Based on the Assumption of Normality.*

3.1.1.1. *The confidence box.* O'Brien et al. (1994) showed that the cost effectiveness plane can be used to present the confidence limits for the estimate of incremental cost effectiveness (Fig. 2).

The following hypotheses are accepted:

1. C_s, C_N, E_s, E_N are distributed according to the normal function.
2. The sample sizes are >30.
3. $\mu_{\Delta C}$ and $\mu_{\Delta E}$ are independent.

The area of the shaded box represents the combined area of confidence for a theoretical example (difference in cost = €15,000, differences in effectiveness = 10

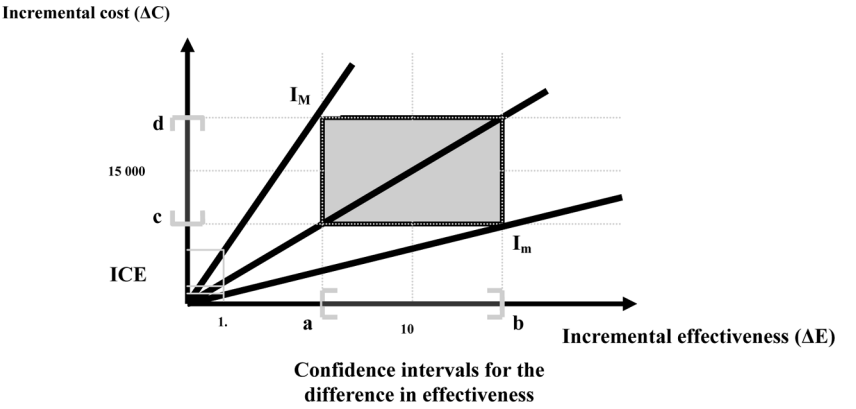

Figure 2. The confidence box.

lives saved). I_m and I_M define, respectively, upper and lower limits for the incremental cost effectiveness ratio (Briggs, 2001).

The principal advantage of this method is its simplicity. One of the principal disadvantages are assumptions about the normality and of the independence of $\mu_{\Delta C}$ and $\mu_{\Delta E}$.

Additionally, under the assumption of independence, the tangents cover less than $(1 - a)100\%$ of the joint probability (Briggs, 2001; Polsky et al., 1997).

A similar approach has been used by Wakker and Klaassen (1995) who have proposed a modified box method to be used when the difference in the effects of two therapies differs significantly from zero. Although the method does not make potentially restrictive assumptions in reference to normality or symmetry, it implicitly assumes the independence of ΔE and ΔC. This method seems to perform better than the original approach in the case when correlation coefficient between ΔE and ΔC is modest (Briggs and Fenn, 1998).

3.1.1.2. *The Taylor theorem.* According to the Taylor theorem, it is the ratio of cost effectiveness that follows a normal distribution well defined.

The approximation of uncertainty by the use of Taylor series has the advantage that we can estimate the covariance between nominator and denominator.

The Taylor formula in the case of ICER (O'Brien et al., 1994) is presented as:

$$\mathrm{Var}(\mathrm{ICER}) \approx \overline{\overline{\mathrm{ICER}}}^2 \left[\frac{\mathrm{Var}(\Delta\mu_C)}{\overline{\overline{\Delta}}^2_{\mu_C}} + \frac{\mathrm{Var}(\Delta\mu_E)}{\overline{\overline{\Delta}}^2_{\mu_E}} + 2\frac{\mathrm{Cov}(\Delta\mu_C, \Delta\mu_E)}{\overline{\overline{\Delta}}_{\mu_C}\overline{\overline{\Delta}}_{\mu_E}} \right]; \qquad (14)$$

or, according to the definition of the coefficient of variation $\mathrm{CV}(X)$ of a random variable X, and of the linear correlation coefficient $\rho(X, Y)$ of two random variables X and Y:

$$\mathrm{Var}(\mathrm{ICER}) \approx \overline{\overline{\mathrm{ICER}}}^2 \left[\mathrm{CV}(\Delta\mu_C)^2 + \mathrm{CV}(\Delta\mu_E)^2 + 2\rho\mathrm{CV}(\Delta\mu_C)\mathrm{CV}(\Delta\mu_E) \right]. \qquad (15)$$

Then, the knowledge of the variance of the ICER estimator permits to determine quite classically the confidence interval of the ICER.

This method appears the more promising in any cases of large samples (>100) very close to the normal distribution (LASS, 2003) with modest correlation coefficient between costs and effects. Nevertheless, the assumption of the normality of the ratio constitutes the mean weakness of the method, particularly in the case where the sample size is not very large. Additionally, we do always take into account the probability that the denominator has a value zero.

3.1.1.3. *The confidence ellipse.* Van Hout and Gordon (1994) argued that the joint cost and density function might be elliptical in shape (Fig. 3), that is, the costs and effects follow a joint normal distribution. They assume that the density function of the couple (Δ_{MC}, Δ_{ME}) has the expression:

$$f(\Delta\mu_C, \Delta\mu_E) = \frac{1}{2\pi\sigma_{\Delta\mu_C}\sigma_{\Delta\mu_E}\sqrt{1-\rho^2}}\exp(Q) \tag{16}$$

where $\sigma_{\Delta\mu_C}$ and $\sigma_{\Delta\mu_E}$ are, respectively, the standard deviations of Δ_{MC} and Δ_{ME}, ρ is the correlation coefficient between $\mu_{\Delta C}$ and $\mu_{\Delta E}$, Q is defined by:

$$Q = \frac{1}{2(1-\rho^2)}\left[\frac{(\mu_{\Delta_C} - \bar{\mu}_{\Delta_C})^2}{\sigma^2_{\Delta\mu_C}} + \frac{(\mu_{\Delta_E} - \bar{\mu}_{\Delta_E})^2}{\sigma^2_{\Delta\mu_E}} - \frac{2\rho(\mu_{\Delta_C} - \bar{\mu}_{\Delta_C})(\mu_{\Delta_E} - \bar{\mu}_{\Delta_E})}{\sigma^2_{\Delta\mu_C}\sigma^2_{\Delta\mu_E}}\right], \tag{17}$$

where $\bar{\mu}_{\Delta_C}$ and $\bar{\mu}_{\Delta_E}$ are, respectively, the means of $\Delta_{\mu C}$ and $\Delta_{\mu E}$. Q = constant that defines the equation of an ellipse centred at $(\bar{\mu}_{\Delta_C}, \bar{\mu}_{\Delta_E})$.

Then it is sufficient to draw the ellipse in such a way as the probability of points belonging to its interior as equal to 0.95 and to draw the tangents issues from the origin I'_m and I'_M. The interval $[I'_m; I'_M]$ can be considered as a confidence interval of the ICER. Seldom are the studies that evaluate the performance of this method, probably because of the relative complexity of obtaining the confidence limits. An internal report of LASS (Laboratoire d'Analyse des Systèmes de Santé) concludes that the ellipse method is the second most performant after the Taylor method for sample size = 100 for each arm and for distributions of costs and effects very close to the normal distribution. On the contrary, Briggs and Fenn (1998) concluded that this method outperforms the Taylor method, indeed when cost and effect differences are almost independent and the coefficient of variation of the denominator is small.

The assumption of joint normality in this case is the principal weakness of the method. The reason is that healthcare costs follow often skewed distribution.

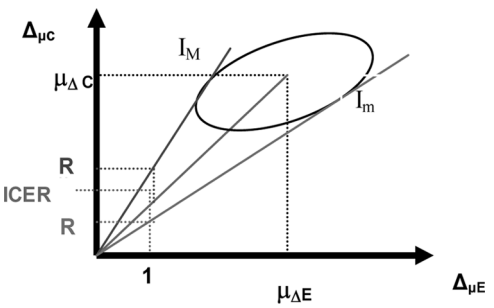

Figure 3. Confidence ellipse on the CE plan.

3.1.1.4. *The Fieller's theorem.* The Fieller theorem covers a general situation in that the nominator and the denominator are dependent with a covariance $\neq 0$. Then, by dividing with the standard error:

$$\frac{\bar{\mu}_{\Delta_C} - (\text{ICER})\bar{\mu}_{\Delta_E}}{\sqrt{\sigma_{\Delta\mu_C} + (\text{ICER})^2\sigma_{\Delta\mu_E} - 2(\text{ICER})\text{cov}(\Delta_{\mu_E}\Delta_{\mu_C})}} \sim N(0, 1). \tag{18}$$

By setting this expression equal to $Z_{1-\frac{\alpha}{2}}$ and by rearranging and solving for ICER this formula leads to define confidence intervals for the ratio as:

$$\begin{aligned}(\overline{\overline{\text{ICER}}})&\frac{1 - z^2_{\alpha/2}\rho\text{cv}(\Delta_{\mu_C})\text{cv}(\Delta_{\mu_E})}{1 - z^2_{\alpha.2}\text{cv}(\Delta_{\mu_E})^2} \\ \pm(\overline{\overline{\text{ICER}}})&\frac{1 - z_{\alpha/2}\sqrt{\begin{array}{c}\text{cv}(\Delta_{\mu_E})^2 + \text{cv}(\Delta_{\mu_E}) - 2\rho\text{cv}(\Delta_{\mu_C})\text{cv}(\Delta_{\mu_E}) \\ -z^2_{\alpha/2}(1-\rho^2)\text{cv}(\Delta_{\mu_E})^2\text{cv}(\Delta_{\mu_E})^2\end{array}}}{1 - z^2_{\alpha/2}\text{cv}(\Delta_{\mu_E})^2}.\end{aligned} \tag{19}$$

The advantage of the Fieller method in reference to the Taylor theorem is that it takes into account the potential asymmetry of the estimator of the ratio and consequently cannot symmetrically be positioned around the point estimation. In the majority of existing researches, Fieller's method outperforms the Taylor method in cost-effectiveness applications and produces reasonably accurate confidence intervals in the majority of the cases considered (Briggs et al., 1999; Heitjan, 2003; Polsky et al., 1997; Tambour and Zethraeus, 1998). Nevertheless, the assumption of joint normality of the sample means, because of the non robustness of the sample mean to skewed data, constitutes the principal weakness of the method. Additionally, if the ICER denominator isn't statistically significant (Heitjan, 2003) this method can produce confidence intervals that they should not be used (i.e., the exterior of an interval).

3.1.2. *The Nonparametric and Bayesian Methods.*

3.1.2.1. *Nonparametric bootstrap.* This bootstrap is particularly valuable when trying to obtain an interval estimate for a ratio of a non symmetric distribution (Good and Hardin, 2006).

This method includes the following stages:

1. Simple sampling with replacement from each treated group.
2. Count $\overline{\overline{\text{ICER}}} = \frac{\bar{\bar{\mu}}_{\text{CN}} - \bar{\bar{\mu}}_{\text{CS}}}{\bar{\bar{\mu}}_{EN} - \bar{\bar{\mu}}_{ES}} = \frac{\bar{\bar{\mu}}_{\Delta C}}{\bar{\bar{\mu}}_{\Delta E}}$ for these new samples.
3. Repeat this operation B times and we will obtain B independent values of $\overline{\overline{\text{ICER}}}$: $\overline{\overline{\text{ICER}}}_1 \ldots \overline{\overline{\text{ICER}}}_B$ the histogram of which is an empirical estimation of the distribution of the ICER estimator.

If we assume that the sampling distribution of the statistic is normal, the confidence interval is based at the distribution of:

$$\frac{\overline{\overline{\overline{\text{ICER}}}}_B - \overline{\overline{\text{ICER}}}_b}{\sigma_B} \tag{20}$$

where $\overline{\overline{\overline{\text{ICER}}}}_B$ is the estimate of the parameter by the bootstrap sample and

$$\sigma_B = \frac{1}{B-1}\sum_{b=1}^{B}\overline{\overline{\overline{\text{ICER}}}}_B - \overline{\overline{\overline{\text{ICER}}}}_b. \tag{21}$$

It is evident that this method may be seriously misleading if the sampling distribution is not normal (Briggs et al., 1997).

Alternatively (Efron, 1987), this estimation permits the definition of a confidence interval for ICER by the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentile values of the bootstrap distribution (for example, the 2, 5th and the 97, 5th percentile give the limits of a 95% confidence interval). Unfortunately, the percentile method doesn't perform all that well: for example, the coverage probability can stray from the nominal value (Armitrage et al., 2001).

One other method for confidence interval estimation is the studentized method. In this case, each bootstrap replicate of the ICER, is transformed into a standardized variable:

$$t = \frac{\overline{\overline{\overline{\text{ICER}}}}_b - \overline{\overline{\overline{\text{ICER}}}}}{s_B}, \tag{22}$$

where s_B denotes the standard deviation of the bootstrap sample. An estimate of the population variance is required to transform the resultant interval into one about ICER. That is, bootstrap could be nested within a bootstrap (Briggs et al., 1997). However, even if the standard error were based on a small value of B, (<30) the total number of bootstrap samples would become very large, typically 20,000–30,000 (Good and Hardin, 2006).

The bias corrected and accelerated method (Efron, 1987) still uses the ordered set $\overline{\overline{\overline{\text{ICER}}}}_1 \ldots \overline{\overline{\overline{\text{ICER}}}}_B$ but chooses the a_1Bth and a_2Bth largest values for the limits of the interval. These values are defined as:

$$a_1 = \Phi\left(w + \frac{w + z_{\alpha/2}}{1 - a(w + z_{\alpha/2})}\right) \quad \text{and} \quad a_2 = \Phi\left(w + \frac{w + z_{(1-a)/2}}{1 - a(w + z_{(1-a)/2})}\right), \tag{23}$$

where Φ is the standard normal cumulative distribution function and $z_{(\alpha/2)}$ is the $100(\alpha/2)$ point of the standard normal distribution. Two adjustments to the percentiles are incorporated into the equation: w adjusts the sampling distribution for the bias; and the quantity α known as the acceleration adjusts for the skewness of the sampling distribution. If these are both zero, then $\alpha_1 = \alpha$ and $\alpha_2 = 1 - \alpha$; the bias corrected and accelerated method (BCa) method reduces to the percentile method (Davision and Hinkley, 1997; Efron and Tibshirani, 1993). A weakness of this method is that the bias correction adjustment while not employing distributional assumptions concerning the sampling distribution of the ICER itself, does make use of parametric assumptions concerning the distribution of the observed bias (Mooney and Duval, 1993). Besides, the evaluations of this method are not especially favorable: Briggs et al. (1999) by using 480 Monte Carlo experiments (they have sited the underlying data to be log normally distributed), showed that BCa method performs better than nonparametric methods but is outperformed by Fieller's method. Heitjan (2003) found out that coverage probabilities for this method are less satisfactory than the Taylor series and Fieller's method and deteriorates very

rapidly as the true values of the ICER get closer to the vertical axis. Additionally, even with these modifications, the use of the nonparametric bootstrap is not recommended for samples of fewer than 100 observations (Good and Hardin, 2006).

3.1.2.2. *The parametric bootstrap.* Parametric bootstrapping is used to consider more realistic scenarios, most frequently to evaluate the performance of the other methods to computing confidence intervals. This is done by assigning a prior probability distribution to each parameter of interest. Essentially, this is a Bayesian approach, with the model parameters being treated as random variables (Lord et al., 1999). The bootstrap algorithm entails taking a large number (B) of samples of size n_i, with replacement, from the original cost and effectiveness data, for each group, by using a defined distribution. These samples permit a confidence intervals estimation of the ICER. Even when we know something about the form of the population distribution the use of parametric bootstrap to obtain interval estimates of the mean cost and effectiveness provides more accurate answers than nonparametric methods. But, we run the risk of introducing error thought, an inappropriate choice of parametric framework (Good and Hardin, 2006).

In general, under a Bayesian approach an interval estimate for ICER can be constructed given that it lies in a particular quadrant. By this approach we can avoid the problems of mixing ratios that lie on opposite sides of the vertical axis. However, the choice of the prior distribution is important and often controversial because it influences the posterior distribution.

3.2. *The Alternative Decision Methods Based on CE Ratios*

Here, the alternative methods to handle the sampling uncertainty are discussed when the uncertainty covers the more than one quadrant.

3.2.1. *The Net Benefit (NB) Framework.* More recently, a number of researchers (Claxton and Posnett, 1996; Stinnett and Mulahy, 1998; Tambour and Zethraeus, 1998) proposed a rearrangement of the cost-effectiveness decision rule in order to overcome the different problems associated with ICER'S. For example, the cost-effectiveness trade off represented by the ratio of two positive differences is not necessarily equivalent to the trade off represented by an equal ratio of negative differences.

In this case, the rule is rearranged as follows:

$$\text{ICER} = \frac{\mu_{CN} - \mu_{CS}}{\mu_{EN} - \mu_{ES}} = \frac{\mu_{\Delta C}}{\mu_{\Delta E}} \leq R_M \quad 0 \leq R_M \mu_{\Delta E} - \mu_{\Delta C}. \tag{24}$$

Or, under the assumption that $R_M \neq 0$:

$$0 \leq \mu_{\Delta E} - \frac{\mu_{\Delta C}}{R_M}. \tag{25}$$

If we note, respectively, $\overline{\overline{\text{NMB}}}$, $\overline{\overline{\text{NHB}}}$, the point estimations of random quantities situated on the right side of the inequalities (24) and (25), the result is:

$$\overline{\overline{\text{NMB}}} = R_M \bar{\bar{\mu}}_{\Delta_E} - \bar{\bar{\mu}}_{\Delta_C} \tag{26}$$

$$\overline{\overline{\mathrm{NHB}}} = \bar{\bar{\mu}}_{\Delta_E} - \frac{\bar{\bar{\mu}}_{\Delta_C}}{R_M}. \tag{27}$$

The advantage of this formulation is that the calculations of the variances of estimators becomes simple. In detail, we have:

$$\mathrm{Var}(\mathrm{NMB}) = R_M^2 \mathrm{Var}(\mu_{\Delta_E}) + \mathrm{Var}(\mu_{\Delta_C}) - 2R_M \mathrm{Cov}(\mu_{\Delta_E}, \mu_{\Delta_C}) \tag{28}$$

$$\mathrm{Var}(\mathrm{NHB}) = \mathrm{Var}(\mu_{\Delta_E}) + \frac{\mathrm{Var}(\mu_{\Delta_C})}{R_M} - 2\frac{\mathrm{Cov}(\mu_{\Delta_E}, \mu_{\Delta_C})}{R_M}. \tag{29}$$

A second advantage is that it is possible to prove that the estimators NMB and NHB are approximately normally distributed (Briggs, 2001).

Consequently, it is simple to count a confidence interval of NMB and NHB in function of R_M. It is clear that there are relations between NB and Fieller's theorems. However, as cost data are often highly skewed and as efficacy data may not be normally distributed, it can be problematic to fit suitable parametric models (Hahn and Whitehead, 2003). Löthgren and Zethraeus (2000) proposed two approaches. The first is based on the assumption that the estimator of NB follows a normal distribution, but the cost and effectiveness data are not specifically modeled. The second is based on the bootstrap method and it permits avoiding distributional assumptions. In a Bayesian approach, the net benefit parameter is treated as random variable, and that has a joint probability distribution specified prior to observation of data. The updating of the prior distribution in the light of data, governed by Bayes's theorem, leads to the posterior distribution.

Bayesian inference is based on the posterior distribution. From it can be calculated a 95% Bayesian probability interval for the Net Benefit, given a threshold value R_M. One problem with the net benefit parameter is that it is dependent on R_M, which is usually not known. One solution to this problem is the cost effectiveness acceptability curve (Briggs and Fenn, 1998).

3.2.2. *Acceptability Curves.* Cost Effectiveness Acceptability Curves (CEAC's), summarize the evidence in favour of the intervention being cost effective for all possible values of the R_M. That is, this curve measures the probability that the CE ratio-resulting from a trial is acceptable in comparison to a predefined CE ratio for all positive values of R_M. Under a frequentist perspective (Fig. 4), the acceptability curve can be estimated by using the NB approach or the ICER approach.

The frequentist approach also can be based either on a parametric or a nonparametric method.

1. Frequentist approach under ICER. It gives the probability that ICER will be below the given line price, under parametric:

$$\mathrm{CEAC}(R_M) = \int_{-\infty}^{+\infty} \int_{-\infty}^{R_M \Delta_{\mu E}} f_{\Delta_{\mu C} \Delta_{\mu E}}(\Delta_{\mu C}, \Delta_{\mu E}) d\Delta_{\mu C} d\Delta_{\mu E}, \tag{30}$$

   where $f_{\Delta_{\mu C}\Delta_{\mu E}}(\Delta_{\mu C}, \Delta_{\mu E})$ denotes the bivariate normal distribution of $(\Delta_{\mu C}, \Delta_{\mu E})$, or nonparametric approach, that uses bootstrap replicate of the ICER:

$$\mathrm{CEAC}(R_M) = \frac{1}{B} \sum_{b=1}^{B} I(\overline{\overline{\mathrm{ICER}}}^b < R_M, \Delta_{\mu E}^b > 0) + I(\mathrm{ICER}^b > R_M, \Delta_{\mu E}^b < 0), \tag{31}$$

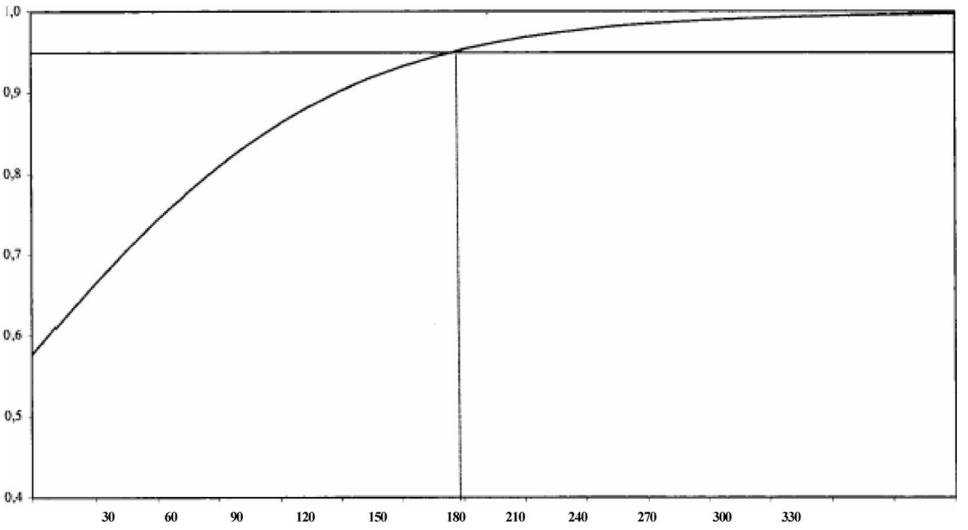

Figure 4. Acceptability curves on CE plane (frequentist approach).

where $I\{A\}$ denotes the standard indicator function; that is, 1 if statement A is true and 0 otherwise.

2. Frequentist approach under Net Benefit. It gives the probability that $\mathrm{NB}(R_M)$ is positive under parametric:

$$\mathrm{CEAC}(R_M) = \int_0^{+\infty} f_{\overline{\overline{\mathrm{NB}}}(R_M)}\left[\overline{\overline{\mathrm{NB}}}(R_M) d\overline{\overline{\mathrm{NB}}}(R_M)\right], \tag{32}$$

or nonparametric approach, that uses, respectively, the B bootstrap replicates of the NB statistic:

$$\mathrm{CEAC}(R_M) = \frac{1}{B}\sum_{b=1}^{B} I\{\overline{\overline{\mathrm{NB}}}^{b}(R_M) > 0\} = 1 - F(0). \tag{33}$$

Under a Bayesian framework, CEAC's give the probability that an intervention is cost effective, under ICER or NB approach.

It has been argued that (Briggs, 1999; Briggs and Fenn, 1998; O'Hagan et al., 2000; Sendi and Briggs, 2001) a Bayesian approach provides a more intuitive and natural framework for decision making. The prior distribution provides a framework for incorporating into the analysis information which is additional to the data from the clinical trial. The main disadvantage is always that it requires the specification of prior distributions for unknown parameters. However, acceptability curves can be used by assuming an uninformative prior distribution but, we must underline that, the precise form of the prior distribution stays important for the small sample sizes.

The advantage of cost-effectiveness acceptability curves over the approach of confidence intervals around ICERs is that they unambiguously quantify the probability an intervention is cost-effective for different R_M and they directly address the study question . The fundamental problem is that important information about the size of the programme is being lost by using a one-dimensional measure

of outcome, the R_M, to summarize a two-dimensional object, namely the joint distribution of incremental costs and effects (Sendi and Briggs, 2001).

4. Discussion

The methods of handling uncertainty are in plain evolution during recent years. However, very little guidance is given to analysts on exactly how this should be done and how the results of any analysis of sampling uncertainty should be presented.

In the majority of existing researches (but not for the entirety), Fieller's and bootstrap methods outperform the Taylor method the confidence ellipse and the box method and produce reasonably accurate confidence intervals. It seems that the form of the data as well as the degree and the direction of the correlation coefficient between differences in costs and effects play an important role to the performance of these methods. It is evident that normal theory methods are inappropriate and produce misleading conclusions even when the data are highly skewed.

On the other hand, the fact that a procedure is non-parametric does not guarantee that it will be robust (O'Hagan et al., 2000). Nonparametric bootstrap will fail if the sample size is small which is frequently the case in randomized clinical studies. Parametric bootstrap can be more accurate if we know something about the population distribution in advance. Nevertheless, we run the risk of introducing error by an inappropriate choice of parametric framework (Good and Hardin, 2006).

Alternative decision models, such as Net Benefit approach, give statistical inferences less problematic than inferences on ICER. One problem with the net benefit parameter is that it is dependent on R_M, which is usually not known. One solution to this problem is the cost effectiveness acceptability curve that summarize the evidence in favor of the intervention being cost effective for all possible values of the R_M. It has been argued that (Bala and Mauskoff, 1999; Briggs, 1999; Briggs and Fenn, 1998; Heitjan et al., 1999; Jones, 1996; O'Hagan et al., 2000; Sendi and Briggs, 2001) a Bayesian approach provides a more intuitive and natural framework for decision making.

In general, Bayesian approaches are considered as most appropriate in the context of economic evaluation by a significant number of authors (Bala and Mauskoff, 1999; Briggs, 1999; Jones, 1996; O'Hagan et al., 2000).

The problem of subjectivity concerning the choice of prior distribution can be handled by the choice of a non informative prior. However, we must underline that, the precise form of the prior distribution stays important for the small sample sizes.

The conclusion that can be drawn from the above is that the most significant factor is to formulate each time the appropriate models on the data and the information that are available to us. Simple nonparametric methods are not automatically robust and there is always a chance that they maybe proved misleading, in relation to the form of the data and the sample size.

References

Armitrage, P., Berry, G. G., Matthews, J. (2001). *Statistical Methods in Medical Research*. Oxford: Blackwell Scientific Publications.

Armstrong, E. P., Khan, Z. M., Perry, A. S., Perri, L. R. (2000). The cost effectiveness of zanamivir and oseltamivir for influenza treatment. *Formulary* 35(12):979–989.

Auray, J.-P., Duru, G., Lamure, M., Pelk, A. (1991). *Fundamentals of Health Economic Evaluation*. Lyon, France: Alexandre Lacassagne Editions.

Bala, M. V., Mauskoff, J. (1999). The estimation and use of confidence intervals in economic analysis. *Drug Inform. J.* 33:841–848.

Briggs, A. (1999). A Bayesian approach to stochastic cost-effectiveness analysis. *Health Econ.* 8:257–261.

Briggs, A. (2001). Handling uncertainty in economic evaluations and presenting the results. In: Drummond, M., McGuire, A., eds. *Economic Evaluation in Health Care: Merging Theory with Practice*. Oxford: Oxford University Press.

Briggs, A., Fenn, P. (1998). Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Econ.* 7:723–740.

Briggs, A., Wonderling, D., Money, C. (1997). Pulling cost effectiveness analysis up by its bootstraps: A nonparametric approach to confidence interval estimation. *Health Econ.* 6:327–340.

Briggs, A., Mooney, C. Z., Wonderling, D. E. (1999). Constructing confidence intervals for costs effectiveness ratios: an evaluation of parametric and nonparametric techniques using Monte Carlo simulation. *Statist. Med.* 18(23):3245–3262.

Claxton, K., Posnett, J. (1996). An economic approach to clinical trial design and research priority setting. *Health Econ.* 5(6):513–524.

Davision, A. C., Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.

Efron, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* 82:171–200.

Efron, B., Tibshirani, R. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Gold, M. R., Siegel, J. E., Russell, L. B., eds. (1996). *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.

Good, P., Hardin, J. (2006). *Common Errors in Statistics (and How to Avoid Them)*. New York: Wiley.

Hahn, S., Whitehead, A. (2003). An illustration of the modelling of cost and efficacy data from a clinical trial. *Statist. Med.* 22:1009–1024.

Heitjan, D. F. (2003). Bayesian methods for analyze cost effectiveness. In: Briggs, A., ed. *Statistical Methods in Medical Research: A Guide to Current Issues and Future Developments*. London: Office of Health Economics (OHE).

Heitjan, D. F., Moskowitz, A. J., Whang, W. (1999). Bayesian estimation of cost-effectiveness ratios from clinical trials. *Health Econ.* 8:191–201.

Jones, D. A. (1996). A Bayesian approach to the economic evaluation of health care technologies. In: Spiker, B., ed. *Quality of Life and Pharmacoeconomics in Clinical Trials*, 2nd ed. Philadelphia, PA: Lippincott-Raven.

LASS (2003). Confidence intervals around ICER's. *Internal Report t*. University Lyon I (UCB).

Lord, J., Asante, M. A. (1999). Estimating uncertainty ranges for costs by the bootstrap procedure combined with probabilistic sensitivity analysis. *Health Econ.* 8(4):323–333.

Löthgren, M., Zethraeus, N. (2000). Definition, interpretation and calculation of cost effectiveness acceptability curves. *Health Econ.* 9:623–630.

Mandelblatt, J. S., Schechter, C. B., Yabroff, K. R., Lawrence, W., Dignam, J., Muennig, P., Chavez, Y., Cullen, J., Fahs, M. (2004). Benefits and costs of interventions to improve breast cancer outcomes in African American women. *J. Clin. Oncol.* 22(13):2554–2566.

Mooney, C., Duval, R. D. (1993). Bootstrapping: a nonparametric approach to statistical inference. *University Paper Series on Quantitative Applications in the Social Sciences*. Series no. 07-095, Sage Newbury Park, CA.

O'Brien, B. J., Drummond, M. F., Labelle, R. J., Willan, A. (1994). In search of power and significance issues in the design and analysis of stochastic cost effectiveness studies in health care. *Med. Care* 32(2):150–163.

O'Hagan, A., Stevens, J., Montmartin, J. (2000). Inference for the cost effectiveness acceptability curve and cost effectiveness ratio. *Pharmacoeconomics* 17(4):339–349.

Pinkerton, S. D., Holtgrave, D. R., Layde, P. M. (2000). Incremental cost-effectiveness of two zidovudine regimens to prevent perinatal HIV transmission in the United States. *Preven. Med.* 30:64–69.

Polsky, D., Glick, H. A., Willke, R., Schulman, K. (1997). Confidence intervals for cost effectiveness ratios: A comparison of four methods. *Health Econ.* 6:243–252.

Pritchard, C. (2004). Developments in economic evaluation in health care: A review of HEED. *OHE Briefing*, Vol. 40. March, OHE Editions.

Sendi, P., Briggs, A. (2001). Affordability and cost effectiveness decision making on the cost effectiveness plane. *Health Econ.* 10:675–680.

Spiegelhalter, D. J., Myles, J. M., Jones, D. R., Abrams, K. R. (2000). Bayesian methods in health technology assessment: A review. *Health Technol. Assess.* 4(38):1–238.

Stinnett, A. A., Mulahy, J. (1998). Net health benefits: a new framework for the analysis of uncertainty in cost effectiveness analysis. *Med. Decis. Making* 18(2 Suppl.): S68–80.

Sophonsritsuk, A., Choktanasiri, W., Weerakiet, S., Rojanasakul, A. (2005). A comparison of outcomes and direct cost between minimal stimulation and conventional protocols on ovarian stimulation in in vitro fertilization. *J. Obstet. Gynaecol. Res.* 31(5):459–463.

Tambour, M., Zethraeus, N. (1998). Bootstrap confidence intervals for cost effectiveness ratios: some simulation results. *Health Econ.* 7(2):143–147.

Van Hout, B., Maiwenn, J., Gilad, S., Gordon, G. S., Rutten, F. (1994). Costs, effects and C/E ratios alongside a clinical trial. *Health Econ.* 3:309–319.

Wakker, P., Klaassen, M. P. (1995). Confidence intervals for cost/effectiveness ratios. *Health Econ.* 4:373–381.

Wykes, T., Reeder, C., Landau, S., Everitt, B., Knapp, M., Patel, A., Romeo, R. (2007). Cognitive remediation therapy in schizophrenia: Randomised controlled trial. *B. J. Psych.* 190:421–427.